

Analisi del riconoscimento delle azioni umane basato su scheletro 3D applicando tecniche di machine learning

Laurea Triennale in Informatica

Tommaso Mattei 1884019

Relatore

Danilo Avola

Correlatore

Daniele Pannone

Anno Accademico 2023/24



SAPIENZA
UNIVERSITÀ DI ROMA



Sommario

- › Contesto
- › Obbiettivi
- › Approcci
- › Soluzione proposta
- › Risultati
- › Conclusioni



Contesto ●○○

Human Action Recognition

- Il riconoscimento delle azioni umane o Human Action Recognition (HAR) si occupa del problema del riconoscimento delle azioni da parte del computer
- Rientra nella branca detta Computer Vision e sfrutta l'intelligenza artificiale per raggiungere il suo scopo
- L'obiettivo di questa branca è riuscire a classificare autonomamente con il machine learning le azioni
- Essa trova applicazioni nell'interazione uomo-macchina, nella medicina e nello sport.



Contesto ○●○

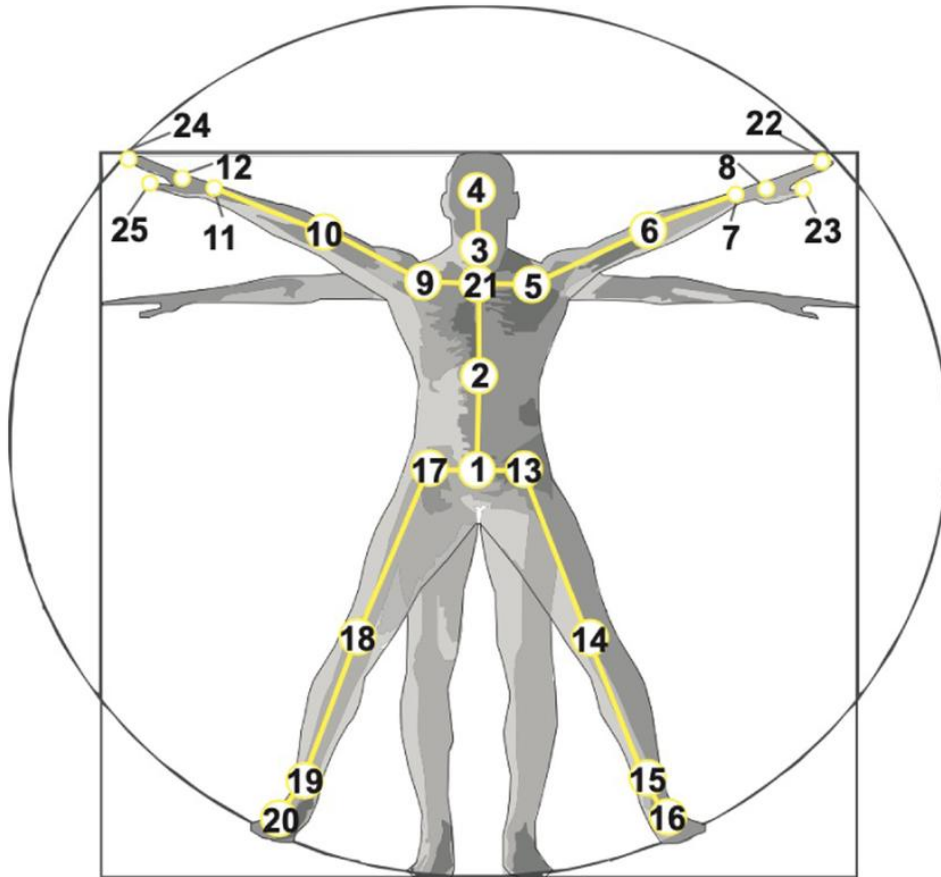
Tipologie di dati

- Nel contesto dell'HAR esistono varie tipologie di dati su cui effettuare la nostra analisi
- Sono presenti: I comuni video a colori RGB, le sequenze scheletriche 2D/3D, mappe di profondità e infrarossi
- Tra queste modalità le più comunemente utilizzate nell'ambito sono I video RGB e le lo scheletro 2D/3D
- Quale tra queste due può portare ad un risultato migliore?



Contesto ○○●

RGB vs Scheletro



- Tra i problemi intrinseci nell'HAR si hanno: la mancanza di illuminazione, la presenza di background complessi e la variazione dei punti di vista
- Il vantaggio principale dei video RGB è la facilità di ottenimento e le informazioni relative agli oggetti e al contesto
- Lo scheletro offre una soluzione ai problemi menzionati e lo sviluppo tecnologico ne ha facilitato l'estrazione a partire dai dati



Sommario

- › Contesto
- › Obbiettivi
- › Approcci
- › Soluzione proposta
- › Risultati
- › Conclusioni



Obbiettivi •

Obiettivo

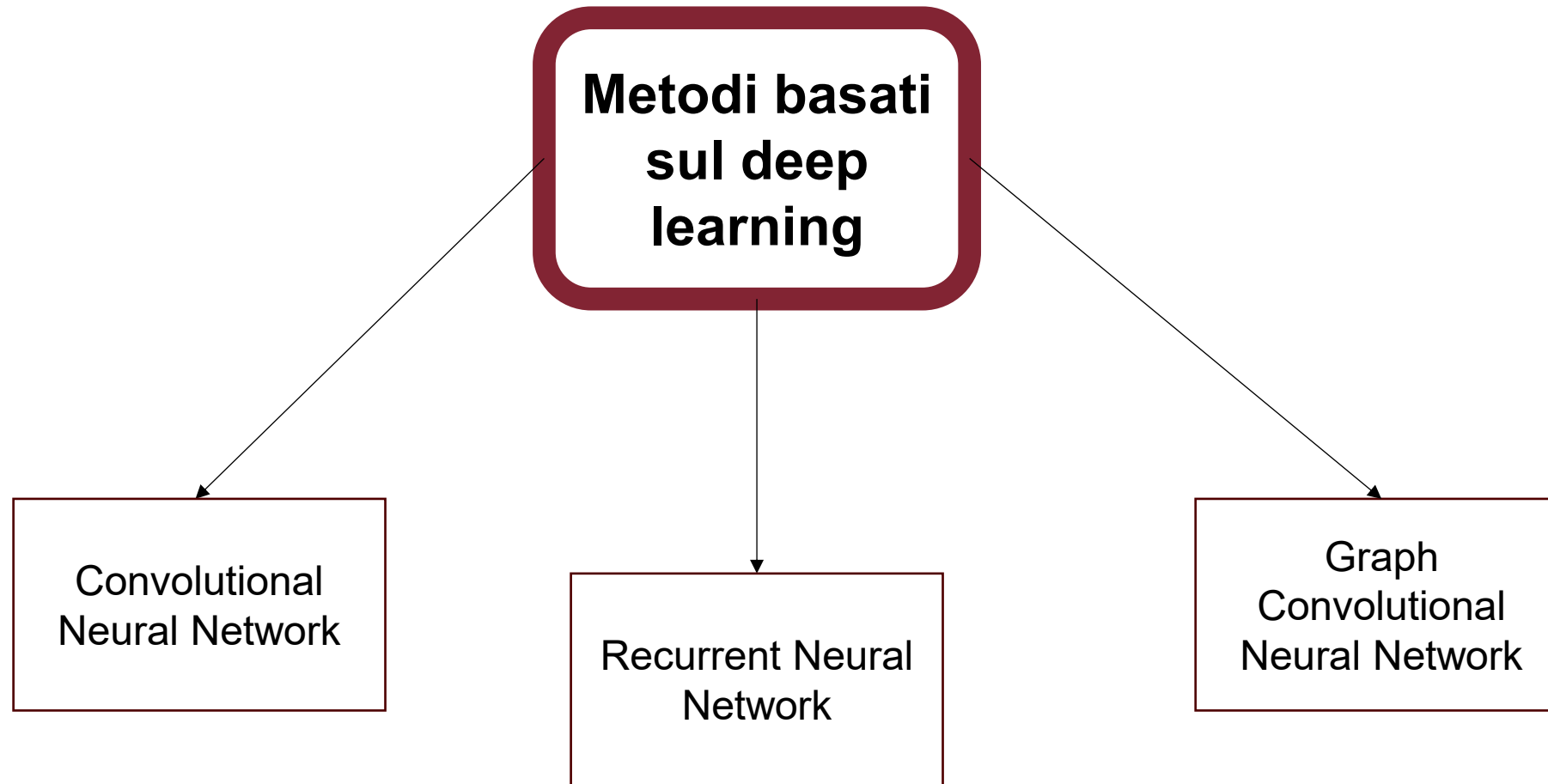
Generalmente parlando, l'obiettivo del Human Action Recognition è prendere in input dei dati e associarvi la corrispondente azione eseguita.

Per quanto riguarda il lavoro di tirocinio, l'obiettivo è stato quello di prendere e modificare uno dei metodi nella letteratura del HAR, per ottenere un'accuratezza finale migliore all'originale.



Sommario

- › Contesto
- › Obbiettivi
- › Approcci
- › Soluzione proposta
- › Risultati
- › Conclusioni





Approcci ○●○

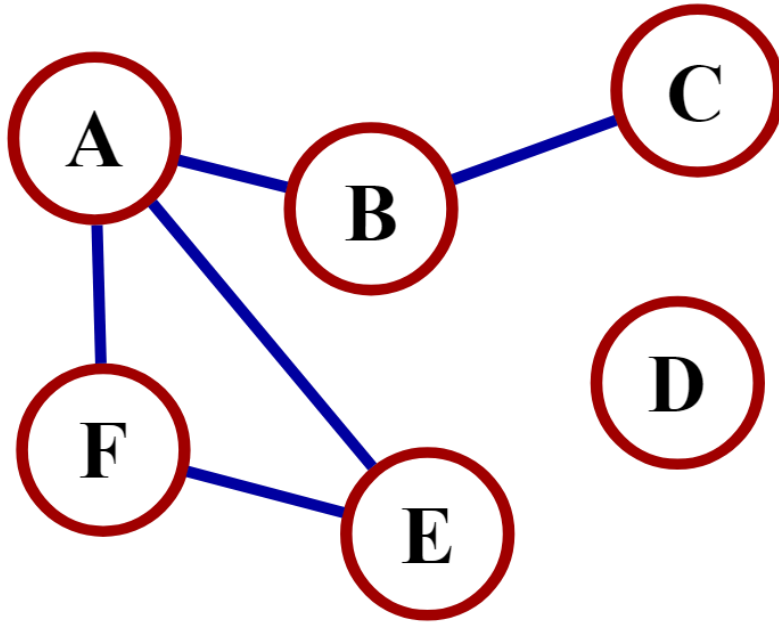
Differenze tra reti neurali

- CNN: l'architettura di questo tipo viene principalmente utilizzata sulle immagini, sfruttando l'operazione della convoluzione sui pixel.
- RNN: la caratteristica principale dell'architettura è il possesso di una memoria, adatta nei problemi in cui il tempo è una componente fondamentale.
- GCN: architettura che lavora modellando ed elaborando le informazioni tramite i grafi. In modo simile ad una CNN sfrutta l'operazione della convoluzione, stavolta non su pixel ma su nodi.



Approcci ○○●

Graph-convolutional Network



- La GCN è l'architettura scelta per questo lavoro
- È facilmente possibile collegare il concetto di grafo con nodi e archi allo scheletro umano con articolazioni e ossa
- La convoluzione tra i nodi del grafo aggrega informazioni tramite un sistema di passaggio di messaggi
- Un singolo nodo possiede informazioni su tutti i propri vicini e oltre



Sommario

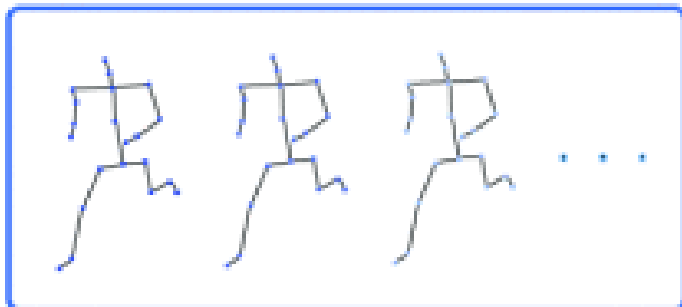
- › Contesto
- › Obbiettivi
- › Approcci
- › Soluzione proposta
- › Risultati
- › Conclusioni



Soluzione Proposta ●○○

Architettura

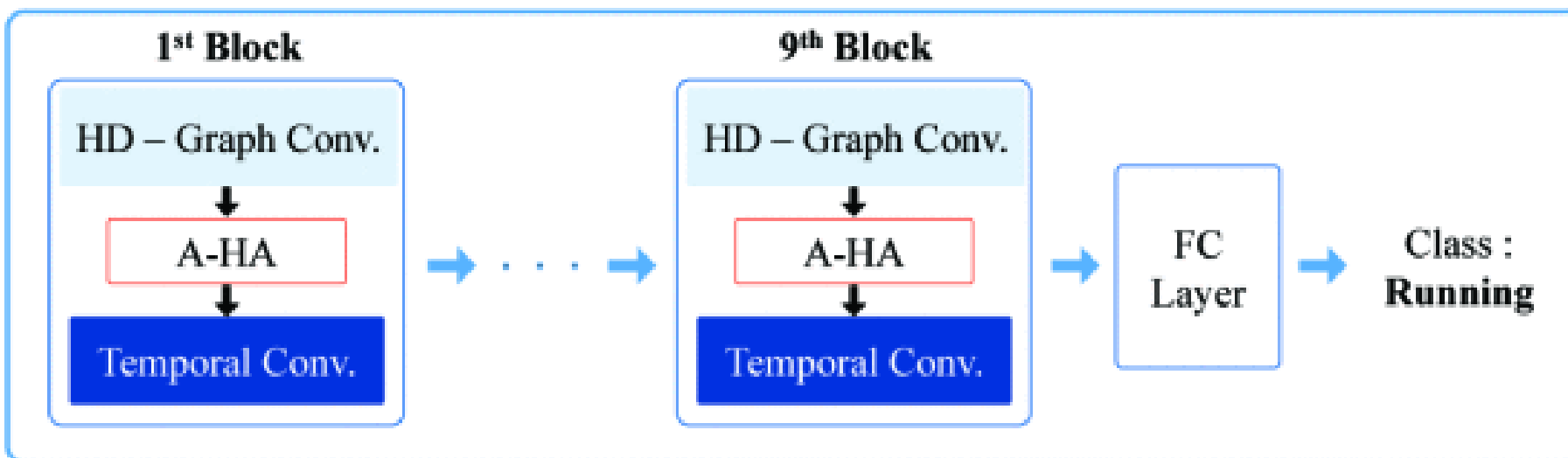
Skeleton Sequence



A – HA

: Attention-Guided
Hierarchy Aggregation

HD – GCN





Soluzione Proposta ○●○

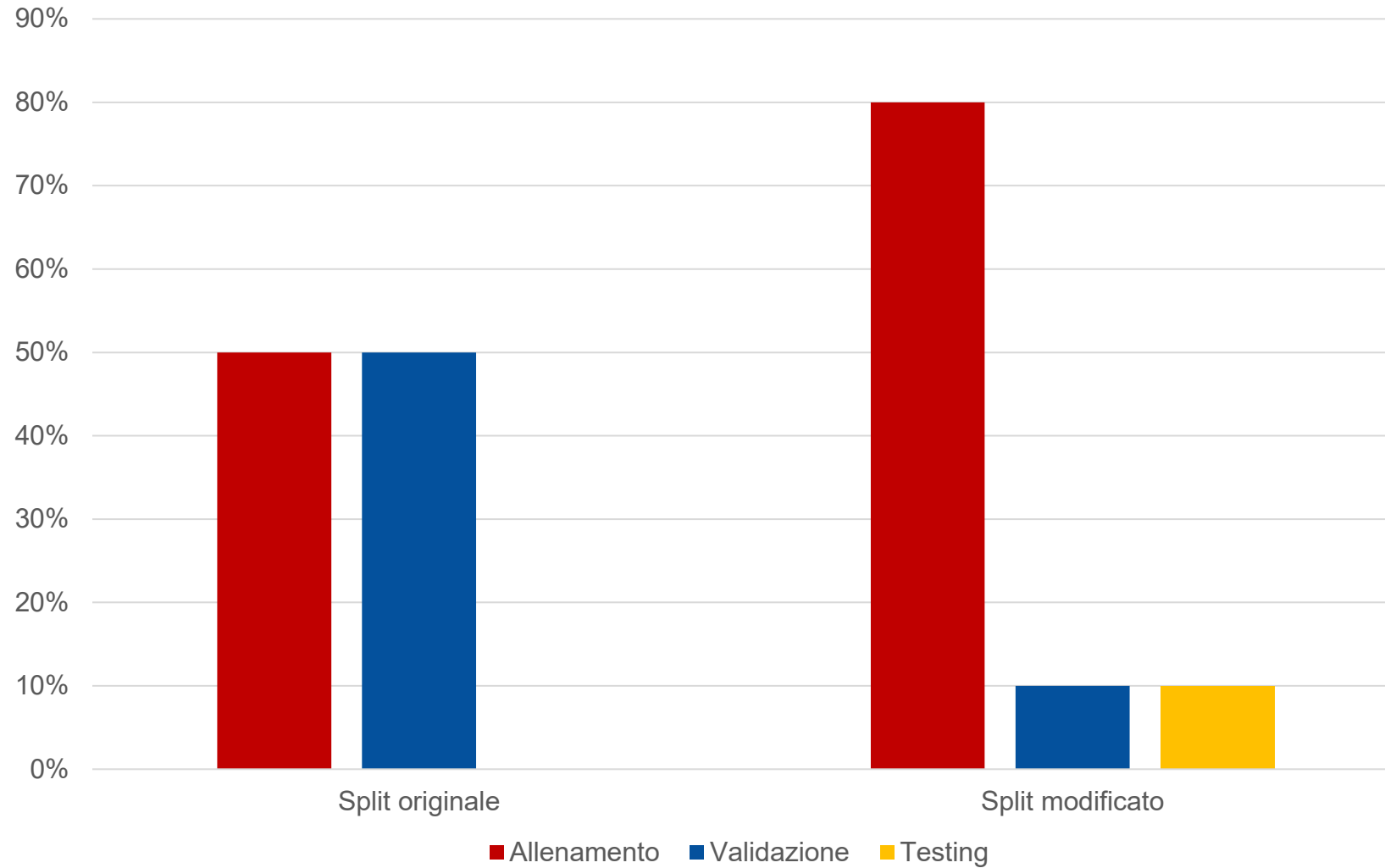
Idee sulle modifiche

- Inizialmente si è pensato di lavorare sulle features del modello, sugli iperparametri oppure su alcuni blocchi dell'architettura stessa
- Analizzando il metodo però si è notato come avvenisse una divisione tra allenamento, validazione e test piuttosto strana
- Con una divisione 50/50 tra allenamento e validation, sfruttando lo stesso insieme di validation per i test finali, si è deciso di intervenire qui



Soluzione Proposta ○○●

Comparazione tra split vecchio e nuovo





Sommario

- › Contesto
- › Obbiettivi
- › Approcci
- › Soluzione proposta
- › Risultati
- › Conclusioni



Risultati ●○○

Dataset

- Il dataset su cui si sono comparati i risultati è l'NTURGB+D 60, specializzato e maggiormente utilizzato per l'HAR relativo allo scheletro 3D
- All'interno del dataset è possibile definire due diverse metriche, scegliendo se analizzare per punti di vista o per soggetti
- I risultati portati sono relativi alla modalità tra soggetti diversi: allenamento, validazione e testing avvengono su individui sempre diversi in ogni fase (Cross Subject)



Risultati ○●○

Accuratezza

Metodo di partenza

93.4%

VS

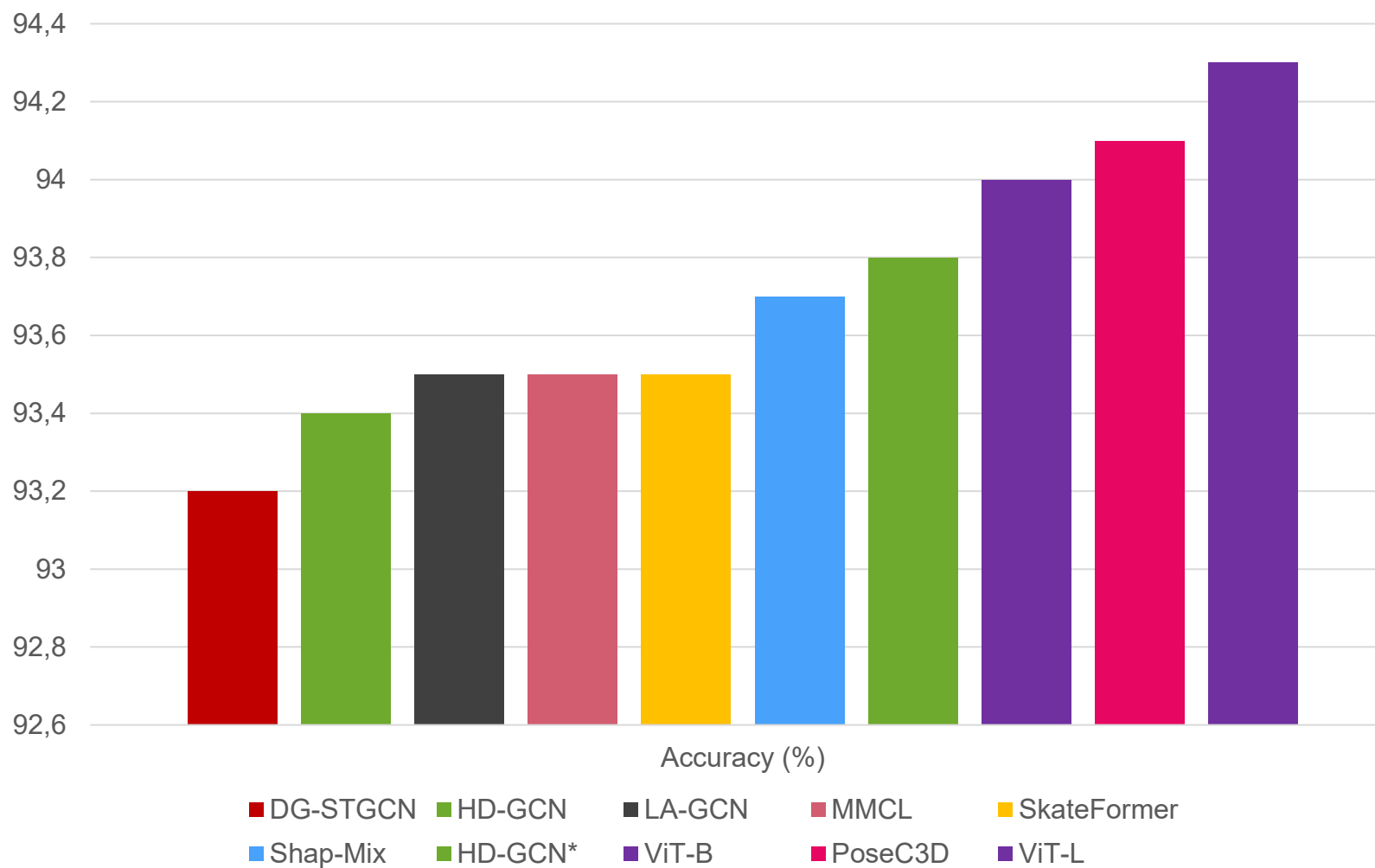
Metodo modificato

93.8%



Risultati ○○●

Confronto





Sommario

- › Contesto
- › Obbiettivi
- › Approcci
- › Soluzione proposta
- › Risultati
- › Conclusioni



Conclusioni •

Conclusioni

Il lavoro svolto ha mostrato dei miglioramenti per l'accuratezza di un modello di deep learning, nell'ambito dello Human Action Recognition.

In futuro sarà possibile ottenere degli ulteriori miglioramenti lavorando sul blocco temporale di una GCN, o come si sta vedendo ultimamente nella letteratura, l'architettura predominante potrebbe diventare quella dei Transformers.



Analisi del riconoscimento delle azioni umane basato su scheletro 3D applicando tecniche di machine learning

*Grazie per aver ascoltato.
Qualche domanda?*